Comparison of 10GigE and InfiniBand Interconnects

Colby Boyer (UC – Berkeley), Joseph Lopez (NMHU), Kimberly Jo McCormick (UNM)

Mentors: Robert Martinez (CTN-5), Todd Bowman (CTN-5),HB Chen (HPC-5), Andree Jacobson (UNM)

Abstract

The goal of a high performance supercomputing cluster is to achieve the highest possible performance using generic parts to the lowest cost. InfiniBand is currently used to maintain low latency and high bandwidth though the recent 10GigE standard offers considerable improvement over its predecessor in both bandwidth and latency performance and might be a viable option. Our driving question is whether or not 10GigE will be price and performance competitive with InfiniBand.

Objectives

Our objective is do a comparison between two cluster interconnect cut through switches: InfiniBand and Fujitsu 10GigE. This will be done by benchmarking the latency performance, throughput performance, cost, and scalability using MPICH2 over InfiniBand over IP and MVAPICH2 over native InfiniBand. We hope our results will show whether 10GigE is a viable interconnect for high speed computing applications.

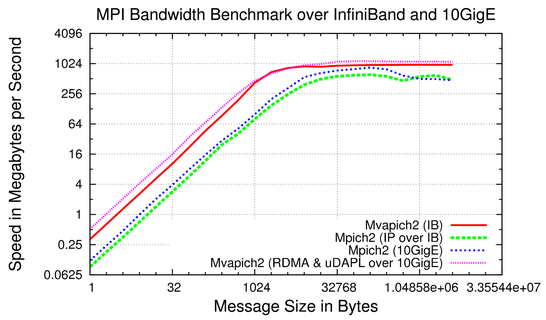
Key Issues

- High performance demands low latency and high bandwidth
- Scalability
  - Expansion of existing networks
  - Maintenance and diagnostics
- Cost Effectiveness

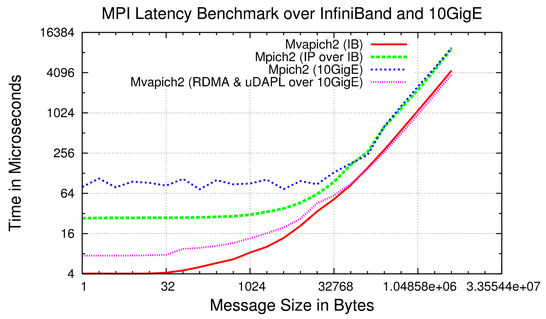
Equipment

- Infiniband 4x
  - Mellanox switch, Voltaire NIC
- 10 GigE
  - Fujitsu switch, NetEffect NIC

Bandwidth Results

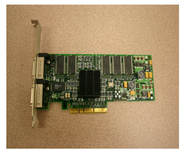


Latency Results



NetEffect 10GigE Card

InfiniBand Card



InfiniBand Cable

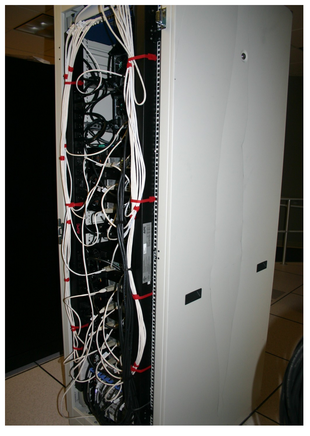
Fujitsu 10GigE Switch

InfiniBand 4x Switch

10GigE Cable

10GigE Cabling

InfiniBand Cabling



OSU Benchmarks (MPICH2, MVAPICH2)

- **Latency**
- **Bandwidth**

Cost Analysis

For a small cluster using copper connections

- 24 Port InfiniBand switch ~$6100 plus cards at ~$500 - $700/ea
  - Per port cost ~$750-$950
- 24 Port 10GigE switch ~$13,000 plus cards at ~ $700/ea
  - Per Port cost ~ $1250

Conclusions

Performance

- InfiniBand achieved lower latency with small message size and comparable with large message size
- 10GigE achieved comparable peak bandwidth
- RDMA (Remote Direct Memory Access) is key to achieving high performance

Evaluation

- InfiniBand has a clear advantage over 10GigE cost wise
- 10GigE can integrate easily into existing TCP/IP networks

Future Research

- Investigate the effects of 10GigE w/ uDAPL on production software programs at LANL
- Increased Scale of Testing
  - Use of multiple swithches (Large Clusters)
  - Integrate with external network systems
- Further cost analysis as prices will change